



**NUMERICAL CALCULATION METHOD,
NUMERICAL CALCULATOR AND
NUMERICAL CALCULATION PROGRAM**

BACKGROUND OF THE INVENTION

The present invention relates to a numerical calculation technique using a computer.

In technique to control a physical quantity such as a pressure or a temperature, successive approximation algorithm has been conventionally used. A residual cutting method is known as highly convergent successive approximation algorithm suitably used for rapidly and accurately obtaining a solution of each parameter to be controlled. Furthermore, Japanese Laid-Open Patent Publication No. 2001-134304 has already described technique to further improve the convergence in the residual cutting method.

In the successive approximation algorithm such as the residual cutting method, the accuracy of a solution is improved by adding a correction quantity to an approximate solution. A sequence of past correction quantities is used for calculating the correction quantity, and the accuracy of the correction quantity can be improved as the sequence is longer. On the other hand, however, as the sequence of past correction quantities is longer, a memory region necessary for the calculation is larger and time required for obtaining the correction quantity is longer.

SUMMARY OF THE INVENTION

An object of the invention is, in numerical calculation using the successive approximation algorithm, suppressing a memory region necessary for the calculation and shortening calculation time while keeping calculation accuracy.

Specifically, the present invention provides a numerical calculation method, a numerical calculator or a recording medium that stores a numerical calculation program for a physical quantity U by solving $A \cdot U = f$, wherein A is a coefficient matrix (in N rows by N columns; wherein N is a positive integer) obtained through discrete of a partial differential equation to be satisfied by the physical quantity U , and f is an inhomogeneous term (source term). In the numerical calculation method, the numerical calculator or the recording medium, the processes of setting 0 as an initial value of a number m of repeating times, giving 0 as an initial value of a perturbation quantity ϕ and setting $(f - A \cdot U^0)$ as an initial value r^0 of a residual r ; and repeatedly executing a first step and a second step while incrementing the number m of repeating times until an approximate solution U^m is converged, and the first step includes the steps of setting an initial value U^0 of the physical quantity U ; obtaining a predicted approximate value ψ^m of $A \cdot \phi = r^m$ through repeated calculation performed by a first calculation unit including an internal solver, and the second step includes the steps of: obtaining, from the predicted approximate value ψ^m , a corrected approximate value ϕ^m for minimizing L^2 norm of a residual r^m through an optimization routine performed by a second calculation unit; and giving $(U^m + \phi^m)$ as an approximate solution U^{m+1} and giving $(r^m - A \cdot \phi^m)$ as a residual r^{m+1} , and in the second step, obtained elements of a vector sequence $A \cdot \phi^m$ are sampled by a given sampling method to be stored in a memory, and a residual minimization coefficient α_l^m (wherein $l = 1, \dots, L$) used for obtaining the corrected approximate value ϕ^m is approximately obtained by using elements of a vector sequence $A \cdot \phi^k$ (wherein $k = m - L + 1, \dots, m - 1$) stored in the memory.

In sampling of elements b_1, b_2, \dots and b_N of the vector sequence $A \cdot \phi^m$ performed in the second step, elements b_i (wherein $i \in \Omega$) are preferably selected, whereas a subset Ω is defined as follows:

$$\Omega = \{i : \text{mod}[i, \lg] = 1\} \cup \{i : |f_i/a_{ii}| > \beta\}$$

wherein \lg is an integer, β is a real number, f_i is an element of the source term and a_{ii} is a diagonal term on the i th row in the i th column of the matrix A .

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a flowchart of a numerical calculation method using the residual cutting method;

FIGS. 2A and 2B are conceptual diagrams of sampling of elements of a vector sequence according to an embodiment of the invention;

10

FIGS. 3A and 3B are diagrams for showing evaluation results of the invention; and

FIG. 4 is a block diagram for conceptually showing the architecture of a numerical calculator of the embodiment.

15

DETAILED DESCRIPTION OF THE INVENTION

A preferred embodiment of the invention will now be described with reference to the accompanying drawings.

20

FIG. 1 is a flowchart of a numerical calculation method according to the embodiment of the invention. FIG. 4 is a block diagram for showing the architecture of a numerical calculator 1 according to the embodiment.

(Formulation of residual cutting equation)

25

It is herein assumed that a physical quantity to be obtained is U , that a coefficient matrix (in N rows by N columns, wherein N is a positive integer) obtained through discrete of a partial differential equation to be satisfied by the physical quantity U is A and that an inhomogeneous term (source term) is f . Under these conditions, an equation to be solved

is:

$$A \cdot U = f \quad \dots \quad (1)$$

This equation can be generally solved by using successive approximation such as the SOR method or the ADI method.

5 In the present invention, the following solution method is employed:

A solution U^∞ of the equation (1) is represented by using an approximate solution U and a perturbation quantity (namely, a difference from a true solution) ϕ as follows:

$$U^\infty = U + \phi \quad \dots \quad (2)$$

10 In the solution method of this invention, the true solution U^∞ of the equation (1) is obtained through correction of the approximate solution U by using the obtained perturbation quantity ϕ .

At this point, a residual r of the approximate solution U is defined as follows:

$$r = f - A \cdot U \quad \dots \quad (3)$$

On the basis of the equations (1) through (3), the following is obtained:

15 $A \cdot (U + \phi) = f$

$$\therefore A \phi = f - AU$$

$$= r$$

Accordingly, in order to obtain the perturbation quantity ϕ , the following equation is solved:

20 $A \phi = r \quad \dots \quad (4)$

(Algorithm for residual cutting method)

In the algorithm employed in the invention, a convergence solution of the equation (4) is not to be obtained but a predicted approximate value ψ is obtained through a minimum unit of repeated calculations with the steepest convergence gradient by the
25 ADI method or the like, and the thus obtained predicted approximate value ψ is supplied to

an optimization control routine. The calculation of the predicted approximate value ψ is repeatedly executed until the result of the optimization control routine satisfies a predetermined condition.

<Optimization control routine>

- 5 Assuming that the number of repeating times is m , a synthesized perturbation quantity ϕ^m for minimizing the L^2 norm of the residual and a new approximate solution U^{m+1} are defined as follows:

$$\phi^m = \alpha_1^m \psi^m + \sum_{l=2}^L \alpha_l^m \phi^{m-l+1} \quad \dots \quad (5)$$

$$U^{m+1} = U^m + \phi^m \quad \dots \quad (6)$$

- 10 wherein α_l^m ($l = 1, 2, 3, \dots, L$) is a residual minimization coefficient and is a constant obtained through calculation described later. The residual is minimized so as to reduce the L^2 norm of the residual as follows:

- When the new approximate solution U^{m+1} is represented by the equation (6), a residual r^{m+1} of this approximate solution is represented as follows on the basis of the
15 equation (3):

$$r^{m+1} = f - A \cdot U^{m+1} \quad \dots \quad (7)$$

When the equation (7) is substituted in the equation (6) and the equations (3) and (5) are used, the following is obtained:

$$r^{m+1} = r^m - \alpha_1^m A \cdot \psi^m - \sum_{l=2}^L \alpha_l^m A \cdot \phi^{m-l+1} \quad \dots \quad (8)$$

- 20 The L^2 norm $\|r^{m+1}\|$ of the residual r^{m+1} of the approximate solution U^{m+1} is given by the following equation (9) as a square root of a sum of all points of $(r^{m+1})^2$:

$$\|r^{m+1}\| = \sqrt{\left(\sum_i^N (r^m - \alpha_1^m A \cdot \psi^m - \sum_{l=2}^L \alpha_l^m A \cdot \phi^{m-l+1})^2 \right)} \quad \dots \quad (9)$$

For minimizing the L^2 norm $\|r^{m+1}\|$ of the residual r^{m+1} , the residual minimization coefficient α_l^m ($l = 1$ through L) of the equation (9) is determined by the least square method as follows:

$$\partial / \partial \alpha_l^m (\|r^{m+1}\|^2) = 0$$

$$l = 1, 2, 3, \dots, L \quad \dots \quad (10)$$

The equations (10) are simultaneous equations of L unknowns with respect to the coefficient α_l^m , and when these equations are numerically solved, the coefficient α_l^m can be defined. When the coefficient α_l^m is defined, the synthesize perturbation quantity ϕ^m is obtained on the basis of the equation (5) and the approximate solution U^{m+1} is obtained on the basis of the equation (6). The calculation method for the residual minimization coefficient α_l^m will be described later.

While incrementing the number m , similar routines are repeated, so as to attain convergence of the solution U^∞ for making zero or minimizing the L^2 norm of the residual of the equation (9).

Now, the numerical calculation method according to this embodiment will be described with reference to the flowchart shown in FIG. 1. This numerical calculation method can be practiced by allowing a computer to execute a program for realizing this numerical calculation method that is recorded in a recording medium such as a CD-ROM 2.

First, in step S1, an initial value U^0 of the physical quantity U to be obtained is set. Then, in step S2, 0 (zero) is given as the initial value of the number m of repeating times, 0 (zero) is given as the initial value of the perturbation quantity ϕ , and the initial value r^0 of the residual r is set to $(f - AU^0)$.

Thereafter, procedures of steps S3 through S7 are repeatedly executed with the number m of repeating times incremented until the approximate solution U^m is converged.

In steps S3 through S6, an approximate solution (predicted approximate value) ψ^m

of the following equation is obtained by using a first calculation unit 10 including an internal solver 11 that executes the successive approximation such as the ADI method, the SOR method or the CG method:

$$A \cdot \phi = r^m$$

5 However, when this equation is to be solved by the existing successive approximation, it is necessary to repeat the calculation too many times for practical use, and hence, the calculation unavoidably takes a massive time. Therefore, in this embodiment, the upper limit N of the number of repeating times of the successive calculation is set (in step S5).

When the predicted approximate value ψ^0 is obtained through steps S3 through S6,
10 this predicted approximate value ψ^0 is used, in step S7, for obtaining a corrected approximate value ϕ^0 for minimizing a next residual r^1 through the optimization routine executed by a second calculation unit 20. On the basis of the equations (2) and (3), this corrected approximate value ϕ^0 is used for obtaining a first-order approximate value U^1 of the physical quantity and a first-order residual r^1 by the following equations:

15
$$U^1 = U^0 + \phi^0 \quad \dots \quad (12)$$

$$r^1 = f - AU^1 \quad \dots \quad (13)$$

When the equation (12) is substituted in the equation (13),

$$\begin{aligned} r^1 &= f - A(U^0 + \phi^0) \\ &= f - AU^0 - A\phi^0 \\ 20 \quad &= r^0 - A\phi^0 \quad \dots \quad (14) \end{aligned}$$

Thus, the first-order residual r^1 can be obtained by the equation (14).

The number m of repeating times is incremented in step S9, and the flow returns to step S3 for repeating similar calculation. When such procedures are repeated, (U^2, r^2) , (U^3, r^3) , ... (U^m, r^m) , ... are successively obtained, and the norm $\|r^m\|$ of the residual is
25 successively reduced.

An equation for obtaining a predicted approximate value ψ^m in step S4 when the number of repeating times is m is as follows:

$$A\phi = r^m$$

Therefore, by using a corrected approximate value ϕ^m optimized through the optimization control routine, the followings are obtained in step S7:

$$U^{m+1} = U^m + \phi^m$$

$$r^{m+1} = r^m - A\phi^m$$

(Method for obtaining residual minimization coefficient α_l)

Herein, δ is introduced as follows for simplification:

$$\delta^m = \psi^m$$

$$\delta^{m-l+1} = \phi^{m-l+1}, \text{ wherein } l = 2, 3, \dots, L$$

In this case, the corrected approximate value ϕ^m is represented as follows:

$$\phi^m = \sum_{l=1}^L \alpha_l^m \delta^{m-l+1}$$

Therefore,

$$r^{m+1} = r^m - A\phi^m$$

$$= r^m - A \sum_{l=1}^L \alpha_l^m \delta^{m-l+1}$$

$$= r^m - \sum_{l=1}^L \alpha_l^m A \delta^{m-l+1}$$

When this equation is expressed without using the symbol Σ ,

$$r^{m+1} = r^m - (\alpha_1^m A \delta^m + \alpha_2^m A \delta^{m-1} + \alpha_3^m A \delta^{m-2} + \dots + \alpha_L^m A \delta^{m-L+1})$$

The residual minimization coefficient α_l^m can be obtained by using, for example, the least square method employing a condition for minimizing the $(m+1)$ th-order residual

norm $\|r^{m+1}\|$ (a square root of a sum of squares).

$$\|r^m\| = \sqrt{\sum_i^N (r^{m+1})^2}$$

$$\|r^m\|^2 = \sum_i^N (r^{m+1})^2$$

$$= \sum_i^N (r^m - \sum_{l=1}^L \alpha_l^m A \delta^{m-l+1})^2$$

5 When this equation is differentiated with respect to α_l^m , the following is obtained:

$$\partial / \partial \alpha_l^m (\|r^{m+1}\|^2) = 0$$

When this equation is solved with respect to α_l^m , the residual minimization coefficient α_l^m can be obtained. The differentiation results in the following:

$$\partial / \partial \alpha_l^m (\|r^{m+1}\|^2) = 2 \sum_i^N \left\{ (r^m - \sum_{l=1}^L \alpha_l^m A \delta^{m-l+1}) \cdot (-A \delta^{m-l+1}) \right\}$$

$$10 \quad \therefore \sum_i^N (A \delta^{m-l+1} \sum_{l=1}^L \alpha_l^m A \delta^{m-l+1}) = \sum_i^N (r^m A \delta^{m-l+1})$$

$$(\delta^m = \psi^m \quad \delta^{m-l+1} = \phi^{m-l+1} \quad (l = 2, 3, \dots, L))$$

wherein $l' = 1, 2, \dots, L$. When these linear simultaneous equations are solved, the residual minimization coefficient α_l^m can be obtained.

For example, when $L = 3$, the above equations are obtained as ternary linear
15 simultaneous equations as follows:

$$\alpha_1^m \Sigma (A \psi^m)^2 + \alpha_2^m \Sigma (A \psi^m A \phi^{m-1}) + \alpha_3^m \Sigma (A \psi^m A \phi^{m-2}) = \Sigma (r^m A \psi^m)$$

$$\alpha_1^m \Sigma (A \phi^{m-1} A \psi^m) + \alpha_2^m \Sigma (A \phi^{m-1})^2 + \alpha_3^m \Sigma (A \phi^{m-1} A \phi^{m-2}) = \Sigma (r^m A \phi^{m-1})$$

$$\alpha_1^m \Sigma (A \phi^{m-2} A \psi^m) + \alpha_2^m \Sigma (A \phi^{m-2} A \phi^{m-1}) + \alpha_3^m \Sigma (A \phi^{m-2})^2 = \Sigma (r^m A \phi^{m-2})$$

Therefore,

$$20 \quad \phi^m = \alpha_1^m \psi^m + \alpha_2^m \phi^{m-1} + \alpha_3^m \phi^{m-2}$$

$$\begin{aligned}
r^{m+1} &= r^m - (\alpha_1^m A \psi^m + \alpha_2^m A \phi^{m-1} + \alpha_3^m A \phi^{m-2}) \\
&= r^m - A \phi^m \\
U^{m+1} &= U^m + \phi^m
\end{aligned}$$

<Sampling of vector sequence>

5 As described above, it is necessary to solve the aforementioned linear simultaneous equations to obtain the residual minimization coefficient α_l^m . In these equations, a vector sequence of $A \cdot \phi^{m-1}$, $A \cdot \phi^{m-2}$, etc. appears frequently. Therefore, in this embodiment, the obtained elements of the vector sequence $A \cdot \phi^m$ are stored in a memory 30 in step S7. Thus, the residual minimization coefficient α_l^m ($l = 1, \dots, L$) to be
10 used for obtaining the corrected approximate value ϕ^m can be obtained by using the elements of the vector sequence $A \cdot \phi^k$ ($k = m - L + 1, \dots, m - 1$) already stored in the memory 30, and hence, the efficiency in the calculation can be improved.

Furthermore, in this embodiment, a method for storing the vector sequence $A \cdot \phi^m$ devised to save the memory capacity and the calculation time will be proposed.
15 Specifically, not all the elements b_1, b_2, \dots , and b_N of the vector sequence $A \cdot \phi^m$ are stored in the memory 30 but part of the elements are sampled to be stored.

At this point, a subset Ω of a set $\{1, 2, \dots, N\}$ is defined as follows:

$$\Omega = \{i : \text{mod}[i, \lg] = 1\} \cup \{i : |f_i/a_{ii}| > \beta\}$$

wherein \lg is an integer parameter, β is a real parameter, a_{ii} is an element on the i th row in
20 the i th column, namely, a diagonal term, of the matrix A , f_i is an element of the source term f . When $\lg = 1$ or $\beta = 0$, $\Omega \equiv \{1, 2, \dots, N\}$. In storing the elements of the vector sequence $A \cdot \phi^m$, not all the elements b_1, b_2, \dots , and b_N are stored but merely elements whose row numbers i are contained in the subset Ω , namely, merely elements b_i ($i \in \Omega$), are selected to be stored. In other words, the aforementioned inner product calculation is
25 approximately performed by using a sum of the sampled elements alone.

FIGS. 2A and 2B are diagrams for conceptually showing exemplified sampling of the elements of the vector sequence $A \cdot \phi^m$. FIG. 2A is a graph for showing the distribution of the respective elements f_i of the source term f , which is normalized by the diagonal term a_{ii} of the matrix A . FIG. 2B shows the exemplified sampling performed on the distribution shown in FIG. 2A, wherein an example A shows a sequence obtained when all the elements are selected, an example B shows a sequence obtained when elements at predetermined intervals (every five elements) are selected, and an example C shows a sequence obtained when the elements belonging to the subset Ω are selected as in this embodiment. Herein, the value of the real parameter β is set on the basis of the maximum value of $|f_i/a_{ii}|$, and $\beta = 0.98 \max(|f_i/a_{ii}|)$.

As is understood from FIGS. 2A and 2B, when the elements belonging to the subset Ω are selected as in this embodiment (namely, in the example C), sampled elements are obtained as a combination of spatially uniformly sampled elements (shown with a symbol \triangle) and so-called locally sampled elements on the basis of the characteristic of the physical phenomenon (shown with symbols \bigcirc and \odot). It goes without saying that the sampling method is not limited to that described herein but spatial sampling similar to the example B alone may be employed or local sampling on the basis of the physical phenomenon alone may be employed. For example, the subset Ω may be defined as follows:

$$\Omega = \{i : \text{mod}[i, \lg] = 1\} \quad \text{or}$$

$$\Omega = \{i : |f_i/a_{ii}| > \beta\}$$

However, in order to save the memory capacity and shorten the calculation time while keeping the calculation accuracy, the combination of the spatial sampling and the local sampling on the basis of the physical phenomenon is preferably employed.

<Evaluation of the invention>

In order to evaluate the effect of the invention, numerical calculation of common equations is actually executed on a computer by the residual cutting method separately in the case where the sampling is not performed (referred to as the case 1) and in the case where the sampling is performed by using the above-described subset Ω (referred to as the case 2). In this evaluation, the equations to be solved are linear simultaneous equations obtained through discrete of the Poisson's equation on a three-dimensional lattice by using a secondary accuracy difference, the SOR method is employed in the internal solver 11, and the number of the residual minimization coefficients α_l^m is 15 ($L = 15$). Also, $lg = 7$, $\beta = 0.98 \max (|f_i/a_{ij}|)$.

FIGS. 3A and 3B show the evaluation results. In FIG. 3B, the ordinate indicates the relative residual $||r||/||f||$, and the abscissa indicates CPU time. It is understood from FIGS. 3A and 3B that when the sampling according to this embodiment is employed, the calculation time is shortened by approximately 15% and the used memory capacity is reduced by approximately 25% as compared with the case where the sampling is not performed.

The numerical calculation herein described is usable in a variety of fields. For example, it is applicable to analysis technology such as fluid analysis as well as to control technology such as driving control for a vehicle.

As described so far, according to this invention, the residual minimization coefficient α_l^m ($l = 1, \dots, L$) to be used for obtaining the corrected approximate value ϕ^m can be approximately obtained by using the elements of the vector sequence $A \cdot \phi^k$ ($k = m - L + 1, \dots, m - 1$) stored in the memory, and hence, the efficiency in the calculation can be improved. In addition, the elements of the vector sequence $A \cdot \phi^k$ are sampled to be stored in the memory, the memory capacity necessary for the calculation can be suppressed and the calculation time can be shortened.